






RESEARCH ARTICLE

WILEY

Urban environment and mortality differentials in Spain

Mathias Voigt¹  | Dariya Ordanovich^{1,2,3}  | Francisco Viciana Fernández⁴  |
 Laura Ann Cilek¹  | Rosa Canovas Balboa⁴ | Diego Ramiro Fariñas¹ 

¹ Institute for Economy, Geography and Demography, Center for Humanities and Social Sciences (CCHS-CSIC), Madrid, Spain

² Department of Production and Dissemination of Demographic and Social Statistics, ESRI Spain, Madrid, Spain

³ Center for Human Geography, Faculty of Geography and History, Complutense University of Madrid (UCM), Madrid, Spain

⁴ Dissemination of Demographic and Social Statistics, Institute of Statistics and Cartography of Andalusia, Seville, Spain

Correspondence

Mathias Voigt, Center for Humanities and Social Sciences (CCHS-CSIC), Madrid, Spain.
 Email: mathias.voigt@cchs.csic.es

Funding information

H2020 Marie Skłodowska-Curie Actions, Grant/Award Numbers: 676060 and CSO2015-69834-R

Abstract

Analyses of health and mortality disparities between today's urban and rural populations appear to be exclusively focused on vastly urbanising countries. By incorporating environmental data at census tract level and accounting for within-area homogeneity, this work attempts to extend classic rural–urban comparisons. Geographical information is linked to a register-based mortality follow up and Spanish census data for the autonomous community of Andalusia. We then apply mixed effects Cox proportional hazards models to estimate individual mortality differences and account for area variations between residential areas. Estimated effects suggest that the shared degree of “urbanicity” does not affect individual hazards of mortality, whereas environmental- and population-based measures influence the relative risk of dying despite controlling for individual-level risk factors. Although we do not find an impact of physical urban measures, our results reveal persistent that area-related mortality disparities which can help to explain the mechanisms behind prevalent spatial-temporal inequalities such as those in Andalusia.

KEYWORDS

mixed effects survival model, small area differences, urban environment, urban population, urbanicity

1 | INTRODUCTION

According to the latest United Nations report on urbanisation, the proportion of the world population living in urban areas is expected to increase from 55% in 2018 to 68% by the year 2050 (United Nations, 2018). Accelerated growth of cities predominantly in developing countries will trigger various changes in social structure, occupational activities, and distribution of wealth. Such a development also entails great challenges regarding social equity and long-term development of public health (Allender, Foster, Hutchinson, & Arambepola, 2008; Woods, 2003).

Recent public health research suggests that various unfavourable health conditions including obesity, high cholesterol, and different

forms of mental illnesses are more prevalent among individuals living in cities, when compared with their rural counterparts (Eckert & Kohler, 2014; Jones-Smith & Popkin, 2010; Van de Poel, O'Donnell, & Van Doorslaer, 2012). The relationship between health, mortality, and environmental features of residential areas, however, is complex. For example, exposure to environmental hazards might have a time-lagged effect on a person's health, yet residential areas are also constantly changing and developing within different cultural and social frameworks (Rydin et al., 2012). In the context of rural–urban differences, environmental effects could also work in both directions, as living in cities is associated with both harmful and health-promoting features at the same time. For example, the increasing adaption of

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors Population, Space and Place Published by John Wiley & Sons Ltd

unfavourable diets in China's urban population raises concerns about future health disadvantages. However, vaccination rates, hygiene, and access to health care are significantly better in China's urban centres than in most rural areas (Gong et al., 2012).

Traditionally, rural and urban subpopulations are compared by using aggregated data sources and dichotomous measures to distinguish between the rural and urban area types. This, however, fails to capture existing heterogeneity and interactions between environmental area features and the variable of interest, which are often masked by differences in the population composition (Freudenberg, Galea, & Vlahov, 2005). Considering such hidden clustering effects and accounting for positive health aspects of urban life like the closer proximity to hospitals and specially qualified doctors, it may appear as if urban residents are more likely to have a health advantage over their rural counterparts.

However, urban dwellers or those who spend most of their day in a highly urbanised area are exposed to specific environmental risk factors like high levels of air pollution and the lack of access to green space (Guan, Zheng, Chung, & Zhong, 2016; Pearce, Shortt, Rind, & Mitchell, 2016). Numerous studies also link unfavourable health effects with increasing spatial proximity to areas with high exposure to negative environmental features, like natural gas wells or high relative amounts of nanomaterial (Colvin, 2003; Fergusson, 1990; Rabinowitz et al., 2015).

Thus, health and well-being of individuals nested in neighbourhoods or other forms of small areas are influenced by shared exposure to certain environmental stress factors. Such effects may occur independently of their individual characteristics and may be mediated through the degree of "urbanicity," referring to the degree or extent of urban features within residential areas (Vlahov, Galea, & Freudenberg, 2005; Xu & Wang, 2015). Even if the majority of health and survival disparities can be traced to behavioural, socio-economic, or biological differences, accounting for exposure to environmental hazards complements the analysis of individual risk factors (Waller & Gotway, 2004).

To our knowledge, only a handful of studies have addressed disparities between rural and urban populations in large regions of high-income countries, possibly due to much lower growth rates when compared with cities in China and other vastly urbanising countries. Even if these changes are modest, attractive job markets in or near urban centres and the growing demand for service work have led to a continuously changing population composition between rural and urban settings with regard to age distribution, education, and other wealth parameters. Consequently, population movements and area developments lead to environmental changes, which in turn affect the population's health (Green, Subramanian, Vickers, & Dorling, 2015; Morris, Manley, & Sabel, 2018; Riva, Curtis, & Norman, 2011). Although comprehensive health insurance coverage in most high-income countries prevents large-scale health and survival disparities, the examination of indirect environmental risk factors, some of which may be specifically urban or rural, can help to explain and ideally prevent the evolving health and mortality gap between social groups. Particularly in places like Southern Europe, it is necessary to analyse

phenomena like urban heat islands where the asphalt and other artificial surfaces store and accumulate summer heat during the day and thereby create a substantially warmer environment during the night for individuals residing and working in urban centres (Oke, 2011).

In the context of this work, we aim to contribute to the debate on urban–rural differences particularly in the field of small area analysis by estimating the effects of environmental impacts and urban characteristics on individual-level survival over time. First, we introduce the conceptual framework focusing on the measurement of "urbanicity." Second, we explain in more detail the construction of our index to measure the latent concept "urbanicity" and introduce the data infrastructure. Third, we fit a Cox proportional hazards (PH) model with mixed effects to estimate individual-level mortality risks and the effects of exposure to urban environment over time. We account for possible stratum homogeneity by including random area effects. Finally, we compare our results to alternative models and discuss the main findings and limitations.

1.1 | Measuring "urbanicity" and small area environment features

The increasing availability of spatially referenced data, greatly improved computing power, and the development of more advanced statistical approaches helps generate new strategies for comparing rural and urban subpopulations. Classic studies often apply dichotomous indicators to distinguish between urban and rural areas. They commonly rely on a core set of characteristics aggregated at differently specified areas, most commonly based on administrative boundaries (Vlahov et al., 2005).

A dichotomous classification according to administrative boundaries appears to be straightforward but fails to capture part of the between-area variation and relevant urban characteristics related to infrastructure, geographic position, and distribution of active space (Cyril, Oldroyd, & Renzaho, 2013). In fact, cities are often surrounded by heavily populated areas that might not be part of the same administrative unit but are still exposed to similar conditions.

To address calls for a more nuanced approach to the subject matter (McDade & Adair, 2001), Vlahov and Galea (2002) proposed a refined conceptualisation for measuring urban space. The crucial step of their approach was to disentangle the two related and often synonymously used concepts urbanisation and "urbanicity." While they defined urbanisation as a process of growth of cities in terms of area and population over time, they related the term "urbanicity" to a current state of an area that can differ by degrees of certain urban characteristics like the proportion of built in environment. In other words, their concept captures the "nature of urban environments" (Dahly & Adair, 2007). "Urbanicity" is a concept strongly dependent on the regional context and changes over time (Leon, 2008). As such, it is difficult to define. The lack of a universal definition, however, provides the opportunity to propose new measures that can then be compared over time.

2 | MATERIALS AND METHODS

2.1 | Data

The *Base de Datos Longitudinal de Población de Andalucía* (BDLPA) is a comprehensive, longitudinal data infrastructure containing administratively collected information on individuals who were registered in the autonomous community of Andalusia at the time of the Spanish population and housing census of 2001. The around 8.3 million individuals are biannually followed up on deaths and emigration based on data derived from regional population registers. A 10% sample of BDLPA data spells can be accessed through the website of the Institute of Statistics and Cartography of Andalusia.¹ The advantage of the BDLPA is that it can be linked to other administrative data, thus allowing us to geo-reference all subjects from the 10% sample and group them into census tracts. To ensure anonymity of individuals and single households, the information was not accessible below census tract level according to the definition of census of 2001. Contextual geographical information is obtained from the CORINE land-cover raster database (coordination of information on the environment), whereas the cartography unit of the Institute of Statistics and Cartography of Andalusia provided maps for the geographical analysis.

2.2 | Indicator of urban environment

The degree of “urbanicity,” or in other words the degree of how urban an area is, refers to a multidimensional and latent concept. Accounting for the complexity of such a concept, we choose a mix between a theoretical and data-driven approach to construct a multicomponent index that allows us to distinguish between different degrees of urban environments (cf. Dahly & Adair, 2007; Jones-Smith & Popkin, 2010; Rey, Jouglu, Fouillet, & Hmon, 2009; Vlahov et al., 2005). By examining graphical tests and correlation coefficients between all accessible environmental variables related to urban settings, we identify four main contributors to the latent concept “urbanicity,” depicted in Figure 1.

First, we calculate the population density per census tract as a measure of relative crowdedness and standardised observations based on the overall deciles to make them comparable with other scale components. Second, we calculate and add the average coverage with medical service based on estimated service area polygons that represent the distance that can be covered between a health facility and

any point on the map within 30-min driving-to-facility time.² Third, we use satellite imagery data from the CORINE land-cover database to estimate the artificial surface area per census tract for the year 2006.³ Fourth, we obtain and add road density by estimating the total length of line objects (transportation networks) within area units (kilometer of road per square kilometer of land surface).

The weights with which single components enter the index are estimated through maximum likelihood factor analysis incorporating standardised single-component values (cf. Brown, 2014; Jreskog, 1967). Factor weights are represented in Table A1. The resulting index variable is further normalised and centred around zero. A Cronbachs alpha score of 91% suggests a sufficient internal consistency of all indicator components. A graphical quality test of the index is depicted in Figure 2, which displays a choropleth map of Seville, the biggest city, population-wise, in Andalusia (INE, 2017). Scores for the

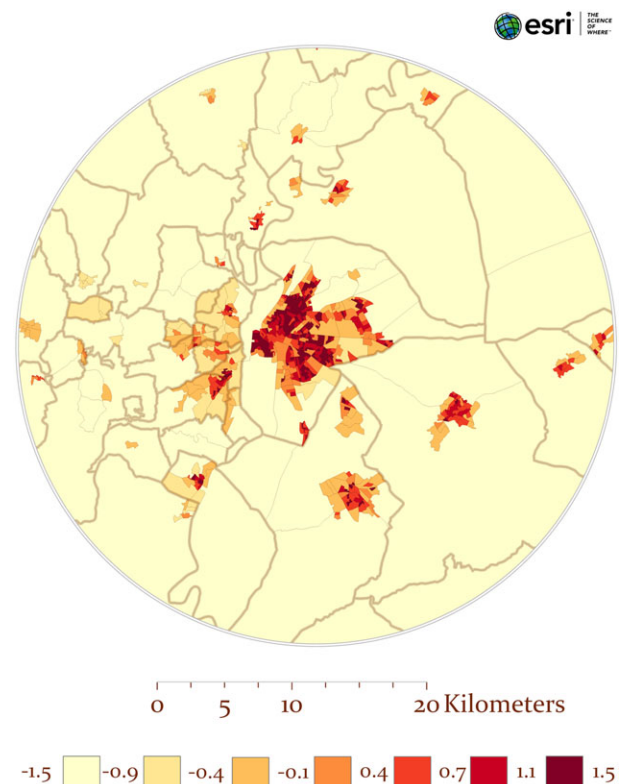


FIGURE 2 Municipality of Seville—Scores for the urban environment indicator by census tract

Indicator components of the “urbanicity” index	
Population Density	Artificial Surface Area
Number of individuals per census tract (standardized and weighted based on the overall deciles)	Proportion of census tract area with artificial surface (population accounted)
Road Density	Proportion of Service Area
Kilometers of roads (transport networks) within census tract (per square km of land surface)	Proportion of census tract area within 30 minutes reach of health facilities (driving to facility time)

FIGURE 1 Components of the “urbanicity” indicator

multicomponent indicator for urban environment are represented by census tract, and darker shades are associated with a higher degree of “urbanicity.” Areas identified as urban are rather small and located in the city centre, which indicates a good graphical fit for our measure.

2.3 | Environmental and population-based area features

The rather physical “urbanicity” indicator does not allow us to capture area-specific heterogeneity regarding environmental hazards and potentially harmful population or social features unrelated to “urbanicity.” As we attempt to distinguish between different kinds of urban and rural areas, it is important to account for further heterogeneity between areas similar to what we can observe in reality. To capture between-area differences regarding the latent socio-economic situation and the exposure to area-specific environmental hazards, we incorporate additional population measures and aggregated survey answers on the residential environment in the analysis. After performing sensitivity tests, we choose perceived average cleanliness, noise exposure, and air pollution to represent additional environmental hazards in our models. The proportion of unemployed individuals at working age and the proportion of single households enter the models as population features. All variables are standardised with reference to the mean values for Andalusia in the year 2001.

2.4 | Study population and individual-level information

There are two main sources for individual-level information. The BDLPA is a mortality follow-up, which is semiannually updated and corrected. By using an individual identifier, we are able to link all subjects in the study to their answers from the population and housing census in 2001, the baseline year for our analysis, and follow them up until 2014. Individuals who died or emigrated within the follow up period are censored at the date of death or emigration. All others are right-censored at the end of the observation period. Information on the residential area and the individual socio-economic context are only available for the baseline year of the study (2001). To reduce the bias induced by potentially unobserved changes in residence and other individual time-varying information, only individuals between the ages 35 and 80 are included in the analysis. The selection of the age groups is premised on informed assumptions about living conditions and tenure status derived from commonly observed life course trajectories in the given age range in Spain (Pla & Cabrerizo, 2004). In general, individuals between ages 35 and 80 have resided in their house or apartment for relatively long periods and are, as they grow older, more likely to own their dwelling. The probability of moving is rather low for those age groups, and individuals are more likely to be exposed to the same environment for the time of our study.

Limiting the age range was further motivated by the distribution of the event of interest as more than 90% of all deaths in Andalusia occur after age 35, but before age 92, the highest age individuals reach at

the end of the follow up period in 2014 (cf. Ocaña Riola & Mayoral-Cortés, 2010; Viciano-Fernández, Ruiz-Ramos, & Pujolar, 2008). A consequence of selecting these age groups is that the sample size decreases from 723,234 to 351,769 individuals. We tested all models with different age ranges but did not observe greatly different patterns in population composition or general model outcomes. To assure that neither the observed population nor the area-specific “urbanicity” and environmental characteristics have dramatically changed over 13 years of observation, particularly in the light of substantial economic fluctuations since 2001 (Escolar-Pujolar, Bacigalupe, & San Sebastian, 2014; Eurostat, 2017), we perform sensitivity tests for different time periods and with different age ranges.

One strength of this analysis lies in the combination of individual-level information with area-specific factors captured in a multilevel setting. Such data structure guarantees that possible index effects are not caused by unobserved heterogeneity between subpopulations. We included sex, disability status, and marital status observed at the baseline year as individual-level variables in our models. To control for socio-economic individual differences, we incorporated several indicators representing social position as for example the highest educational degree, ownership status of the dwelling, and car ownership status. All socio-economic variables are derived from the census questionnaire of 2001.

2.5 | Statistical approach

The incorporation of area effects into an analysis of individual-level mortality differences requires statistical testing for potential impacts of cluster-specific effects and, in case of geographical data, the spatial distribution. The graphical representation of the multicomponent “urbanicity” index and statistical tests regarding the environmental variables suggest that observations are more likely to be similar if they are geographically closer to each other. To assess if observations are spatially autocorrelated, intrinsic stationarity is assumed before we calculate a row-standardised matrix of spatial weights that is based on the list of contiguous neighbours. At least one point of the boundary of a spatial polygon that represents a census tract has to be within snap distance of at least one point of a neighbouring polygons' border to meet our contiguity condition (cf. Anselin, 2013; Bivand, Pebesma, & Gomez-Rubio, 2013). We then calculate the product-moment correlation coefficient (Moran's I) as statistical test for spatial autocorrelation (Moran, 1948; Sokal & Wartenberg, 1983). Although spatial autocorrelation of mortality indicators would justify the analysis of area differences in the first place, we also estimate Moran's I for other central variables including the index for “urbanicity” and aforementioned area-specific environmental features.

2.6 | Statistical model

Following the descriptive tests for spatial autocorrelation, we estimate mortality disparities by degree of “urbanicity” and environmental impact with an extended version of the Cox PH model. The original

model is the most commonly used approach to model censored time to event data, particularly when the main interest is to obtain relative effects of covariates (Mills, 2011). Such effects are assumed to be proportional over time and enter the model multiplicatively as expressed in the following equation (cf. Kleinbaum & Klein, 2010).

$$h(t) = h_0(t) \exp(\beta_i X_i), \quad (1)$$

where h_0 is the baseline hazard and $\exp(\beta_i X_i)$ is the nonnegative function of covariates. Hazard ratios are obtained through the maximisation of the partial log likelihood with respect to $\beta_i X_i$ (Allison, 2014; Therneau & Grambsch, 2000). Because only the right-hand side of the formula is maximised, the Cox PH model does not require you to specify the underlying baseline distribution. Due to our assumption that individuals are nested in small areas where they are exposed to similar environmental hazards and the same degree of "urbanicity," we choose to apply an extended version of the Cox PH model, which allows us to account for such homogeneity within clusters. Thus, a stratum-specific frailty term is added to the original model (cf. Austin, 2017). The resulting Cox PH model with mixed effects can be considered as shared frailty model with a normally distributed stratum specific frailty term Z_j as follows (Therneau, 2015).

$$h(t_{ij}) = h_0(t) \exp(X_i \beta + Z_j), \quad (2)$$

where Z_j is the design matrix for random effects, which captures homogeneity within clusters. The model can be interpreted as multilevel survival model with shared frailties. The added random effect term can be understood as relative effect of given covariate patterns on the baseline hazard, which varies across census tracts (Pankratz, De Andrade, & Therneau, 2005). Given the set-up of our analysis, it is necessary to account for left truncation (Cain et al., 2011). This adjustment affects survival estimates for everybody in the sample because their time under risk of dying before the start date of the study remains unobserved. In other words, we select individuals based on their survival upon the start year of the examination. To account for left-truncation and assure we measure age-specific mortality differences, we choose to use person years as the time scale in our models. Cohort effects are accounted for by including birth cohort effects as covariate (Canchola et al., 2003).

3 | RESULTS

In order to analyse area differences, spatial variation must be present in the variable of interest. We determine to what extent these differences are spatially associated by estimating Moran's I for the variables in our analysis, which can be interpreted as the correlation between a variable and its spatial neighbours (Anselin, Syabri, & Smirnov, 2002). These estimates and associated p values are presented in Table 1. All variables of interest exhibit significant spatial autocorrelation, justifying our analysis of area differences in Andalusia.

As the analysis aims to highlight the impact of shared environmental factors on individual-level survival, the population under

TABLE 1 Moran's I Statistics for univariate relationships of indicators with geographic dimension

	Moran's I statistic	Test statistic p value
Smoothed SMR	0.380	<0.001
"Urbanicity" indicator	0.784	<0.001
Perceived neighbourhood pollution	0.444	<0.001
% Unemployment	0.500	<0.001

Note. SMR: standardized mortality ratio.

observation is considered to be nested in geographical units, which requires the application of a multilevel model structure. The estimated coefficients (fixed effects) of four separate Cox PH models with mixed effects and a step-wise increasing number of covariates are presented in Tables 2 and 3. When compared using likelihood ratio tests, these four presented models fit significantly better than a Cox PH model without random effects. Furthermore, as these models include more explanatory variables, fit improves.

The estimated hazard ratio in Model 1 indicates that alone, the selected urban features do not substantially affect individual mortality differences between census tracts in Andalusia. In the second model, we account for within-area population heterogeneity by incorporating socio-economic and demographic individual-level variables with well-documented indirect effects on mortality. The estimates for individual-level variables, depicted in Table 3, show typical mortality patterns. Men between ages 35 and 92 have a substantially higher relative mortality risk compared with their female counterparts. Estimates further suggest that individuals with functional limitations and other disabilities have a mortality hazard more than three times higher than those without these impediments. Moreover, all socio-economic variables point towards increased relative risks for less wealthy and less educated individuals, with reference to both the more affluent and those with university education. As the estimated hazards for these individual-level risk factors change only marginally with the incorporation of environmental and population features, we present them in a separate table to avoid distraction from the effects of primary interest.

After including individual-level differences (Model 2), the effect of urban environment appears more pronounced than in Model 1. In Model 2, every unit increase in the "urbanicity" scale of a census tract increases the estimated hazard of dying by three percentage points. Although changes in the individual-level impact factors are negligible between different models, the effect of the degree of "urbanicity" on survival varies with the incorporation of additional area-level characteristics. In Model 3, we incorporate the effects of perceived cleanliness, noise, air, and water pollution in an attempt to control for different kinds of heterogeneity between urban areas with the same degree of "urbanicity." The estimates suggest that including such environmental area features reduces the effect of the degree of "urbanicity" to 1.7 percentage points for every unit increase. Both cleanliness and pollution are found to have a highly significant but small effect on survival. Perceived cleanliness of the area is estimated

TABLE 2 Cox proportional hazards model with mixed effects—Second level fixed environmental effects on individual hazards

	Hazard ratios (95% CI)			
	Model 1	Model 2	Model 3	Model 4
“Urbanicity” index	1.0061 (0.9963, 1.0158)	1.0308*** (1.0212, 1.0404)	1.0176** (1.0058, 1.0295)	1.0027 (0.9904, 1.0150)
Perceived cleanness			1.0007** (1.0001, 1.0012)	1.0005* (0.9999, 1.0011)
Perceived pollution			1.0020*** (1.0011, 1.0029)	1.0020*** (1.0012, 1.0029)
Perceived noise			0.9999 (0.9991, 1.0008)	0.9995 (0.9987, 1.0004)
% Unemployed				1.0025*** (1.0013, 1.0036)
% Single HH				1.0094*** (1.0070, 1.0119)
Individual variables		X	X	X
Observations	351,769	351,769	351,769	351,769
Integr. log likelihood	−542,763.5	−538,537.7	−538,510.2	−538,471.8
LR test statistic	128.61 (<i>df</i> = 2)***	8,451.6 (<i>df</i> = 12)***	55 (<i>df</i> = 15)**	76.8 (<i>df</i> = 17)***

* $p < 0.01$. ** $p < 0.005$. *** $p < 0.001$.

to increase the hazard by 0.07% for every unit of increase. Results also suggest the hazard increases by 0.2% per every one additional percentage point of perceived pollution. In Model 4, when we include population characteristics of small areas, the estimated hazard for

the “urbanicity” indicator is very close to one. As this “urbanicity” effect shrinks, we find that the proportions of both unemployed individuals and single households in a census tract increase individual hazards by 0.25 and 0.94 percentage points, respectively. The addition of

TABLE 3 Cox proportional hazards model with mixed effects—Individual fixed effects corresponding to models in Table 1

	Hazard ratios (95% CI)		
	Model 2	Model 3	Model 4
Male Reference: female	2.0872*** (2.0682, 2.106)	2.0883*** (2.0962, 2.1074)	2.0898*** (2.0707, 2.1089)
Physically dependent Reference: no dependency	3.0286*** (2.9617, 3.0954)	3.0180*** (2.9512, 3.0848)	3.0028*** (2.9360, 3.0695)
Single	1.4124*** (1.3826, 1.4422)	1.4126*** (1.3829, 1.4424)	1.4038*** (1.3740, 1.4336)
Widowed	1.1836*** (1.1586, 1.2085)	1.1823*** (1.1573, 1.2073)	1.1808*** (1.1559, 1.2058)
Divorced/separated Reference: Married	1.4794*** (1.4249, 1.5340)	1.4765*** (1.4219, 1.5310)	1.4768*** (1.4222, 1.5314)
No or incomplete education	1.3798*** (1.3387, 1.4210)	1.3837*** (1.3425, 1.4249)	1.3942*** (1.3529, 1.4357)
Primary/secondary education Reference: tertiary education	1.1595*** (1.1155, 1.2036)	1.1598*** (1.1157, 1.2038)	1.1652*** (1.1211, 1.2093)
Does not own house/apartment Reference: owns house/apartment	1.1625*** (1.1362, 1.1888)	1.1598*** (1.1157, 1.801)	1.1652*** (1.211, 1.2093)
Does not own car Reference: owns car(s)	1.2690*** (1.2493, 1.2887)	1.2693*** (1.2497, 1.2890)	1.2653*** (1.2456, 1.2851)
Birth cohort	0.9818*** (0.9790, 0.9846)	0.9817*** (0.9789, 0.9846)	0.9817*** (0.9789, 0.9846)
Observations	351,769	351,769	351,769

* $p < 0.01$. ** $p < 0.005$. *** $p < 0.001$.

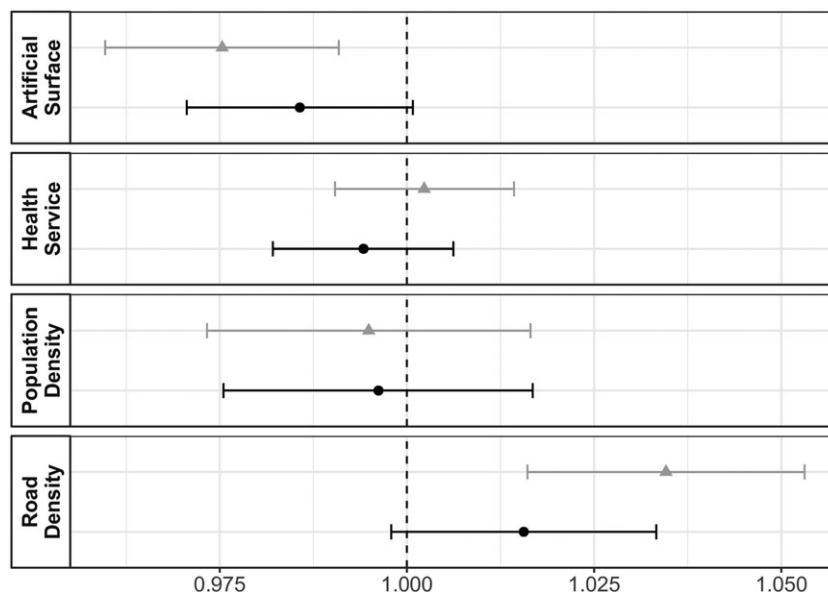


FIGURE 3 Estimated hazard ratios (95% CI) for “urbanicity” indicator components (grey, Model 1, black, Model 4)

population-based measures does not appear to influence the effect of environmental area features.

We also estimate possible effects of single indicator components on the mortality hazards in a given census tract. These results (Figure 3) indicate that road density negatively affects survival in a model with mixed effects and without additional area variables, whereas the percentage of artificial surface has a slightly positive effect on survival. Just as with the index variable, the effect of single components vanishes when accounting for further environmental and social variables. Given that areas with the same level of “urbanicity” are still quite heterogeneous, the presumed negative effects of road density, urban contamination, and other explicitly urban risk factors appear to be less important in explaining mortality disparities between census tracts.

An advantage of shared frailty models over classic survival approaches is the ability to estimate relative effects of covariate

patterns on the baseline hazard across clusters. Assuming subjects are exposed to shared environmental risk factors that, in spite of individual-level differences, influence their survival risk, we incorporate a normally distributed random effect for the residential area in the model. Estimated median frailties and its variations for all models are depicted in Figure 4. Naturally, the random variation is lower in models where we account for the shared additional area effects. Nevertheless, if translated into risk scores, there are still substantial differences between census tracts. For example, in Model 4, a census tract about one standard deviation above the mean corresponds to a relative mortality risk of 1.119. In other words, there is almost a 12 percentage point increase in the hazard of dying compared with the mean census tracts. Further variation measures can be found in Table A2, which we also provide as a summary of the likelihood ratio tests between all models and their counterparts without the additional random effects. The test

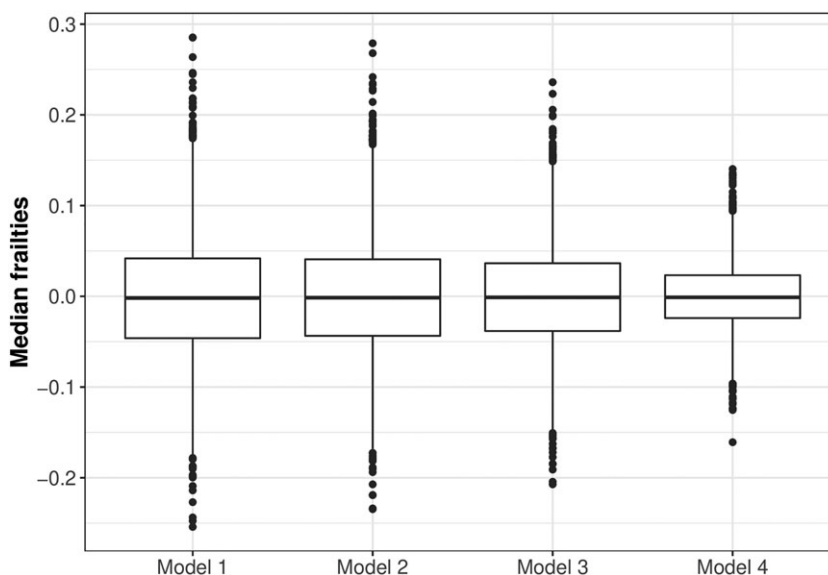


FIGURE 4 Random area effects (median frailty) by model

statistics indicate that models with these shared frailties improve the fit significantly.

Figure 5 shows the effect of shared frailty by census tract. Values are presented in exponentiated form and can therefore be understood as risk scores. The darker the shade, the higher the unexplained relative mortality risk of individuals in the respective area. Although the majority of census tracts experience mortality risks close to the mean (risk scores between 0.97 and 1.03), some hotspots appear to exist in central Andalusia, where mortality risk is up to 15 percentage points higher than average. Contrary to previous analyses of spatio-temporal differences, these high mortality areas do not appear to be clustered in the southwest area of the region but are instead spread randomly throughout Andalusia (Ocaña Riola & Mayoral-Cortés, 2010).

4 | DISCUSSION

In this work, we aim to extend classic approaches for analysing mortality disparities between rural and urban subpopulations. To capture the rural-urban gradient and estimate “urbanicity” effects on survival, we disentangle population and physical features from environmental impact factors in residential areas. We use satellite imagery data and census information to identify four universal predictors of “urbanicity,” the multidimensional latent concept that describes the “urban nature” of an area (Dahly & Adair, 2007). Our index represents an improvement over classic binary measures that are based solely on population density. The use of census tracts as a clustering unit increases the precision with regard to area size and reduces the risk of misclassifying large areas as urban if, for example, only a part of the overall area exhibits urban features. Therefore, our approach offers an advantage over comparable measures, such as the “rurality index” proposed by

Ocaña Riola and Sánchez-Cantalejo (2005) in which data were aggregated at the municipality level.

We incorporate our index in mixed effects Cox PH models to estimate long-term survival according to different degrees of “urbanicity” along with individual and shared environmental risk factors. Results suggest that individuals residing in areas with higher levels of unemployment, single households, and perceived pollution face small but highly significant survival disadvantages, even after controlling for individual-level risk factors. Although population-based and environmental factors appear to explain the majority of geographical survival differences in modern-day Andalusia, we found no clear evidence that physical urban environment, as captured through the aforementioned “urbanicity,” index had an effect on survival. The initial negative impact of “urbanicity” disappeared when incorporating other small area variables into the model, indicating that the physical urban concept may mask effects in other spheres. Although initial models appeared to show a small effect of the more precise physical measures on individual-level survival, the index does not explain small area differences in mortality in Andalusia after controlling for additional information on environmental and social measures. Because our results focus on a single region in Spain, these population and environment effects on survival should be examined in other contexts. Further analysis can highlight potential risk factors in different residential area types and their effect on growing inequalities in individual-level survival.

Some limitations and threats to validity exist due to data availability, unobserved mediators, and assumptions undertaken when conducting the analysis. On the basis of general life course trajectories in the context of Andalusia, we trust that individuals are unlikely to change their residence after age 35 (Pla & Cabrerizo, 2004). However, although residency likely remains stable, other central explanatory variables such as our “urbanicity” indicator and some individual-level

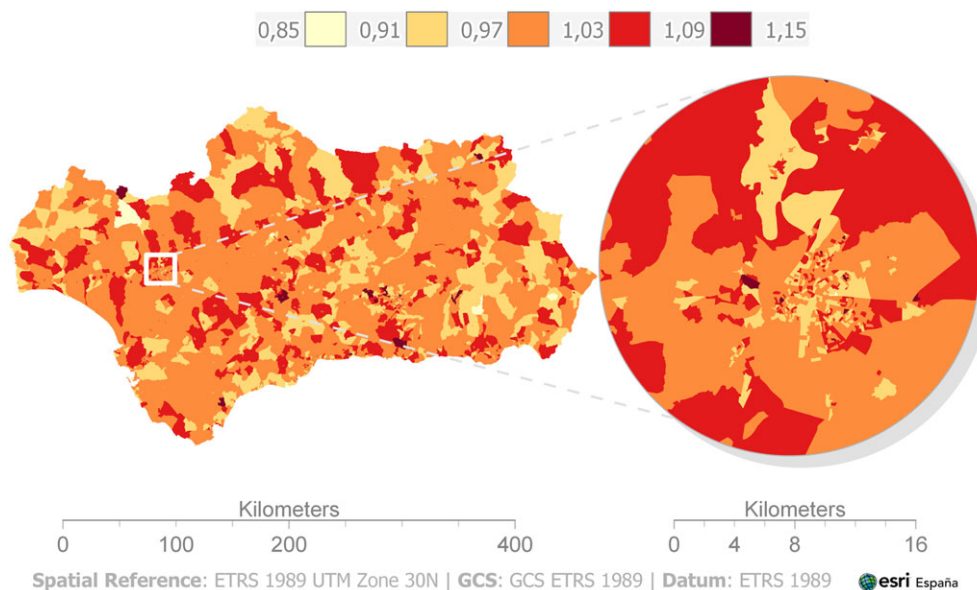


FIGURE 5 Exponentiated random effects by census tract in Andalusia and Seville (zoom-in)

information have almost certainly changed over time. Because we cannot estimate the extent of these variations with the available single cross-sectional data point in 2001, we have to assume environmental and individual indicators like car ownership remain relatively constant over the time period of our study.

The most recent financial and debt crisis of 2008 hit Andalusia particularly hard and led to significant job loss, a continuously increasing at-risk-of-poverty rate, and a high number of evictions (Eurostat, 2017; Romanos, 2014). Thus, the assumption of residential continuity may not hold among economically disadvantaged groups. Moreover, there is no information measuring the average exposure to the estimated environmental and social features of the residential area. The average amount of time someone spends in his or her residential area and is therefore exposed to its environment probably differs by age, employment status, and other unmeasured area features such as access to "third places" (Mehta & Bosson, 2010; Oldenburg & Brissett, 1982).

In spite of these data constraints, to our knowledge, this analysis is the first study to combine detailed small area (census tract) information, individual-level variables, and survival follow up in Southern Europe. Our results differ from a previous analysis on a larger provincial scale for Spain (Regidor et al., 2015). This other study suggested a negative association between per-capita income and average survival times, although we find that environmental features and, to a greater extent, population composition affect survival probability. Moreover, our exploratory survival analysis contributes to the debate on how individual health and socio-economic differences relate to spatially correlated mortality differences. Such research can help to explain the mechanisms behind prevalent spatial-temporal inequalities such as those in Andalusia (Ocaña Riola & Mayoral-Cortés, 2010).

Although our results could not explain geographical differences based on a more detailed measure of urban space, we found that area conditions such as high levels of perceived pollution and a high percentage of unemployed coresidents increase individual relative mortality risks in the presence of other well-known protective individual characteristics. Future research must continue to explore and account for the role area heterogeneity plays in individual mortality.

ACKNOWLEDGEMENT

We thank the Institute of Statistics and Cartography of Andalusia, Spain, for providing generously their data and time. This work was supported by the LONGPOP (Methodologies and Data Mining Techniques for the Analysis of Big Data Based on Longitudinal Population and Epidemiological Registers) project, which has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant 676060 and the project LONGHEALTH (Early Life Conditions and Health and Mortality in Adult Life CSO2015-69834-R) funded by the Spanish Ministry of Economy and Competitiveness.

CONFLICT OF INTEREST

We declare no conflict of interest in conducting this research.

FUNDING INFORMATION

This work was supported by the LONGPOP project, which has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 676060 and the project LONGHEALTH CSO2015-69834-R funded by the Spanish Ministry of Economy and Competitiveness.

ENDNOTES

¹See website: <http://www.juntadeandalucia.es/institutodeestadisticaycartografia/longevidad/index-en.htm>

²The estimation is conducted with the *Network Analyst extension* (ArcMap 10.5). To limit the polygons to inhabited areas, we intersect them with a 250 × 250 m of population grid layer and yield the portion of populated service areas in square kilometres

³According to the classification, artificial surfaces include urban fabrics (continuous and discontinuous urban fabric surfaces), industrial and commercial areas (industrial or commercial units and transport units, road and rail networks and associated land, and port areas and airports), mine and dump sites (mineral extraction sites and construction sites, dump sites, and construction sites), and artificial nonagricultural areas (green urban areas vegetated areas and sport and leisure facilities). The original data are provided at a minimum scale of 1:100000

ORCID

Mathias Voigt  <https://orcid.org/0000-0002-6639-7456>

Dariya Ordanovich  <https://orcid.org/0000-0001-7127-8518>

Francisco Viciano Fernández  <https://orcid.org/0000-0002-2135-3855>

Laura Ann Cilek  <https://orcid.org/0000-0002-0096-6337>

Diego Ramiro Fariñas  <https://orcid.org/0000-0002-2766-307X>

REFERENCES

- Allender, S., Foster, C., Hutchinson, L., & Arambepola, C. (2008). Quantification of urbanization in relation to chronic diseases in developing countries: A systematic review. *Journal of Urban Health*, 85(6), 938–951. <https://doi.org/10.1007/s11524-008-9325-4>
- Allison, P. D. (2014). *Event History and Survival Analysis* (2nd. Ed). Thousand Oaks, CA: SAGE publications. <https://doi.org/10.4135/9781452270029>
- Anselin, L. (2013). *Spatial econometrics: Methods and models* (Vol. 4). Springer, Dordrecht: Springer Science & Business Media. <https://doi.org/10.1007/978-94-015-7799-1>
- Anselin, L., Syabri, I., & Smirnov, O. (2002). Visualizing multivariate spatial correlation with dynamically linked windows. *Urbana*, 51, 61801.
- Austin, P. C. (2017). A tutorial on multilevel survival analysis: Methods, models and applications. *International Statistical Review*, 85(2), 185–203. <https://doi.org/10.1111/insr.12214>
- Bivand, R. S., Pebesma, E., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R* (Vol. 10). New York: Springer Science & Business Media. <https://doi.org/10.1007/978-1-4614-7618-4>
- Brown, T. A. (2014). *Confirmatory Factor Analysis for Applied Research* (2nd. Ed). New York: The Guilford Press.

- Cain, K. C., Harlow, S. D., Little, R. J., Nan, B., Yosef, M., Taffe, J. R., & Elliott, M. R. (2011). Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. *American Journal of Epidemiology*, 173(9), 1078–1084. <https://doi.org/10.1093/aje/kwq481>
- Canchola, A. J., Stewart, S. L., Bernstein, L., West, D. W., Ross, R. K., Deapen, D., Pinder, R., Reynolds, P., Wright, W., and Anton-Culver, H. (2003). Cox regression using different time-scales. Western Users of SAS Software. San Francisco, California.
- Colvin, V. L. (2003). The potential environmental impact of engineered nanomaterials. *Nature Biotechnology*, 21(10), 1166–1170. <https://doi.org/10.1038/nbt875>
- Cyril, S., Oldroyd, J. C., & Renzaho, A. (2013). Urbanisation, urbanicity, and health: A systematic review of the reliability and validity of urbanicity scales. *BMC Public Health*, 13(1), 513. <https://doi.org/10.1186/1471-2458-13-513>
- Dahly, D. L., & Adair, L. S. (2007). Quantifying the urban environment: A scale measure of urbanicity outperforms the urban rural dichotomy. *Social Science & Medicine*, 64(7), 1407–1419. <https://doi.org/10.1016/j.socscimed.2006.11.019>
- Eckert, S., & Kohler, S. (2014). Urbanization and health in developing countries: A systematic review. *World Health & Population*, 15(1), 7–20. <https://doi.org/10.12927/whp.2014.23722>
- Escobar-Pujolar, A., Bacigalupe, A., & San Sebastian, M. (2014). European economic crisis and health inequities: Research challenges in an uncertain scenario. *International Journal for Equity in Health*, 13(1), 59. <https://doi.org/10.1186/s12939-014-0059-5>
- Eurostat (2017). European Union statistics on income and living conditions. Report, European Union.
- Fergusson, J. E. (1990). *Heavy elements: Chemistry, environmental impact and health effects*. Oxford: Pergamon.
- Freudenberg, N., Galea, S., & Vlahov, D. (2005). Beyond urban penalty and urban sprawl: Back to living conditions as the focus of urban health. *Journal of Community Health*, 30(1), 1–11. <https://doi.org/10.1007/s10900-004-6091-4>
- Gong, P., Liang, S., Carlton, E. J., Jiang, Q., Wu, J., Wang, L., & Remais, J. V. (2012). Urbanisation and health in China. *The Lancet*, 379(9818), 843–852. [https://doi.org/10.1016/S0140-6736\(11\)61878-3](https://doi.org/10.1016/S0140-6736(11)61878-3)
- Green, M. A., Subramanian, S. V., Vickers, D., & Dorling, D. (2015). Internal migration, area effects and health: Does where you move to impact upon your health? *Social Science & Medicine*, 136, 27–34.
- Guan, W. J., Zheng, X. Y., Chung, K. F., & Zhong, N. S. (2016). Impact of air pollution on the burden of chronic respiratory diseases in China: Time for urgent action. *The Lancet*, 388(10054), 1939–1951. [https://doi.org/10.1016/S0140-6736\(16\)31597-5](https://doi.org/10.1016/S0140-6736(16)31597-5)
- INE (2017). Series detalladas desde 2002—Resultados por Comunidades Autonomas.
- Jones-Smith, J., & Popkin, B. M. (2010). Understanding community context and adult health changes in China: Development of an urbanicity scale. *Social Science & Medicine*, 71(8), 1436–1446. <https://doi.org/10.1016/j.socscimed.2010.07.027>
- Jreskog, K. G. (1967). A general approach to confirmatory maximum likelihood factor analysis. ETS Research Report Series. *Psychometrika*, 34(2), 183–202.
- Kleinbaum, D. G., & Klein, M. (2010). *Survival analysis* (Vol. 3). New York: Springer.
- Leon, D. A. (2008). Cities, urbanization and health. *International Journal of Epidemiology*, 37(1), 4–8. <https://doi.org/10.1093/ije/dym271>
- McDade, T. W., & Adair, L. S. (2001). Defining the urban in urbanization and health: A factor analysis approach. *Social Science & Medicine*, 53(1), 55–70. [https://doi.org/10.1016/S0277-9536\(00\)00313-0](https://doi.org/10.1016/S0277-9536(00)00313-0)
- Mehta, V., & Bosson, J. K. (2010). Third places and the social life of streets. *Environment and Behavior*, 42(6), 779–805. <https://doi.org/10.1177/0013916509344677>
- Mills, M. (2011). *Introducing survival and event history analysis*. London: Sage Publications. <https://doi.org/10.4135/9781446268360>
- Moran, P. A. P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2), 243–251. <https://doi.org/10.1111/j.2517-6161.1948.tb00012.x>
- Morris, T., Manley, D., & Sabel, C. E. (2018). Residential mobility: Towards progress in mobility health research. *Progress in Human Geography*, 42(1), 112–133. <https://doi.org/10.1177/0309132516649454>
- Ocaña Riola, R., & Mayoral-Cortés, J. M. (2010). Spatio-temporal trends of mortality in small areas of Southern Spain. *BMC Public Health*, 10(1), 26. <https://doi.org/10.1186/1471-2458-10-26>
- Ocaña Riola, R., & Sánchez-Cantalejo, C. (2005). Rurality index for small areas in Spain. *Social Indicators Research*, 73(2), 247–266. <https://doi.org/10.1007/s11205-004-0987-3>
- Oke, T. (2011). Urban heat island. In *The Routledge handbook of urban ecology* (pp. 120–131). Oxon: Routledge Abingdon.
- Oldenburg, R., & Brissett, D. (1982). The third place. *Qualitative Sociology*, 5(4), 265–284. <https://doi.org/10.1007/BF00986754>
- Pankratz, V. S., De Andrade, M., & Therneau, T. M. (2005). Random-effects cox proportional hazards model: General variance components methods for time-to-event data. *Genetic Epidemiology*, 28(2), 97–109. <https://doi.org/10.1002/gepi.20043>
- Pearce, J., Shortt, N., Rind, E., & Mitchell, R. (2016). Life course, green space and health: Incorporating place into life course epidemiology. *International Journal of Environmental Research and Public Health*, 13(3), 331. <https://doi.org/10.3390/ijerph13030331>
- Pla, A. C., & Cabrerizo, J. A. (2004). Home ownership and social inequality in Spain. In *Home ownership and social inequality* (pp. 233–254). Stanford: Stanford University Press.
- Rabinowitz, P. M., Slizovskiy, I. B., Lamers, V., Trufan, S. J., Holford, T. R., Dziura, J. D., ... Stowe, M. H. (2015). Proximity to natural gas wells and reported health status: Results of a household survey in Washington County, Pennsylvania. *Environmental Health Perspectives*, 123(1), 21–26. <https://doi.org/10.1289/ehp.1307732>
- Regidor, E., Vallejo, F., Giraldez-García, C., Ortega, P., Santos, J. M., Astasio, P., & de la Fuente, L. (2015). Low mortality in the poorest areas of Spain: adults residing in provinces with lower per capita income have the lowest mortality. *European Journal of Epidemiology*, 30(8), 637–648. <https://doi.org/10.1007/s10654-015-0013-x>
- Rey, G., Jougl, E., Fouillet, A., & Hmon, D. (2009). Ecological association between a deprivation index and mortality in France over the period 19972001: Variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health*, 9(1), 33. <https://doi.org/10.1186/1471-2458-9-33>
- Riva, M., Curtis, S., & Norman, P. (2011). Residential mobility within England and urban-rural inequalities in mortality. *Social Science & Medicine*, 73(12), 1698–1706. <https://doi.org/10.1016/j.socscimed.2011.09.030>
- Romanos, E. (2014). Evictions, petitions and escraches: Contentious housing in austerity Spain. *Social Movement Studies*, 13(2), 296–302. <https://doi.org/10.1080/14742837.2013.830567>
- Rydin, Y., Bleahu, A., Davies, M., Dvila, J. D., Friel, S., De Grandis, G., ... Wilsson, J. (2012). Shaping cities for health: Complexity and the planning of

- urban environments in the 21st century. *Lancet*, 379(9831), 2079–2108. [https://doi.org/10.1016/S0140-6736\(12\)60435-8](https://doi.org/10.1016/S0140-6736(12)60435-8)
- Sokal, R. R., & Wartenberg, D. E. (1983). A test of spatial autocorrelation analysis using an isolation-by-distance model. *Genetics*, 105(1), 219–237.
- Therneau, T. (2015). Mixed effects cox models. CRAN repository.
- Therneau, T. M. and Grambsch, P. (2000). Extending the cox model. Edited by P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg, page 51.
- United Nations (2018). World urbanization prospects: The 2018 revision. Report, Department of Economic and Social Affairs, Population Division. <https://doi.org/10.18356/02486bd4-en>
- Van de Poel, E., O'Donnell, O., & Van Doorslaer, E. (2012). Is there a health penalty of China's rapid urbanization? *Health Economics*, 21(4), 367–385. <https://doi.org/10.1002/hec.1717>
- Viciana-Fernández, F., Ruiz-Ramos, M., & Pujolar, A. E. (2008). Andalucía y España. In *Desigualdades en la Esperanza de Vida*, book section 3. Asociación para la Defensa de la Sanidad Publica de Andalucía. Madrid: Elivisar Espana.
- Vlahov, D., & Galea, S. (2002). Urbanization, urbanicity, and health. *Journal of Urban Health*, 79, S1–S12. https://doi.org/10.1093/jurban/79.suppl_1.S1
- Vlahov, D., Galea, S., & Freudenberg, N. (2005). The urban health advantage. *Journal of Urban Health*, 82(1), 1–4. <https://doi.org/10.1093/jurban/jti001>
- Waller, L. A., & Gotway, C. A. (2004). *Applied spatial statistics for public health data* (Vol. 368). Hoboken, New Jersey: John Wiley & Sons. <https://doi.org/10.1002/0471662682>
- Woods, R. (2003). Urban-rural mortality differentials: An unresolved debate. *Population and development review*, 29–46.
- Xu, Y., & Wang, F. (2015). Built environment and obesity by urbanicity in the US. *Health & Place*, 34, 19–29. <https://doi.org/10.1016/j.healthplace.2015.03.010>

How to cite this article: Voigt M, Ordanovich D, Viciana Fernández F, Cilek LA, Canovas Balboa R, Ramiro Fariñas D. Urban environment and mortality differentials in Spain. *Popul Space Place*. 2019;e2239. <https://doi.org/10.1002/psp.2239>

APPENDIX A

TABLE A1 Factor loadings and uniqueness parameter from the maximum likelihood factor analysis

	Factor loading	Uniqueness parameter
Population density	0.97	0.05
% of Artificial surface area	0.82	0.33
% of Health service area	0.90	0.19
Road density	0.65	0.58

Note. Factor explains 71% of the variance.

TABLE A2 Random effects (REs) statistics and model comparison with model without random effects

	Model 1	Model 2	Model 3	Model 4
Standard deviation RE	0.1559	0.1251	0.1202	0.1125
Variance RE	0.0243	0.0156	0.0144	0.0127
AIC	1,085,414	1,077,044	1,077,001	1,076,938
Log-likelihood ratio test	127.11***	56.36***	48.36***	37.55***
Chi-square (df)	(1)	(1)	(1)	(1)

Note. AIC, Akaike Information Criterion.

* $p < 0.01$. ** $p < 0.005$. *** $p < 0.001$.